Cite this: DOI: 10.1039/xxxxxxxxxx

Information Processing in Cooperative RNA networks†

Cole Mathis,*[a] SES591 Spring16,[b‡] and Sara Imari Walker[a]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

A major hurdle in understanding the origin of life on Earth lies in understanding how interactions between chemical species can give way to complex life-like phenomena. It is of particular interest to understand how chemical species can *cooperate* in such a way that each species promotes the growth of others. One example of this type of cooperation can be found in the *Azoarcus* ribozyme system[1,2]. Recently information theory has proved useful in analyzing biological systems[3]. Here we use a numerical model of the *Azoarcus* ribozyme system in order to extend that analysis and we use the techniques employed by Kim et. al[3] in an attempt to elaborate on the dynamics of cooperation in the Azoarcus system.

1 Introduction

The origin of life is the greatest unsolved mystery in the history of science. Understanding the origin of life on Earth requires an understanding of how prebiotic chemistry gave way to biology. In that regard, RNA chemistry has provided tractable models of chemical evolution. Accordingly, there have been extensive experimental and theoretical studies on the nature of RNA enzymes or *ribozymes*.

In the context of the origin of life, it is essential to understand how prebiotic chemistry set the stage for Darwinian evolution. Accordingly, there has been and continues to be an exhaustive search for a self replicating RNA molecule. However, to date, a self-replicating ribozyme remains elusive[4,5]. In response, some authors have suggested that the first self-replicating entities may not have been individual molecules but rather communities collectively autocatalytic chemical species[1]. Recently the *Azoarcus* ribozyme system has been studied in this context[1].

Herein I develop a numerical model of the *Azoarcus* ribozyme system. I apply two information measures which elaborate on the dynamics of cooperation in that chemical system. I compare the results of that analysis with the results of a recent study on information processing in biological systems. I find that the information measures do not identify the relevant dynamics for understanding cooperation in this system. I will discuss this and some implications for understanding biological evolution as a subset of chemical evolution.

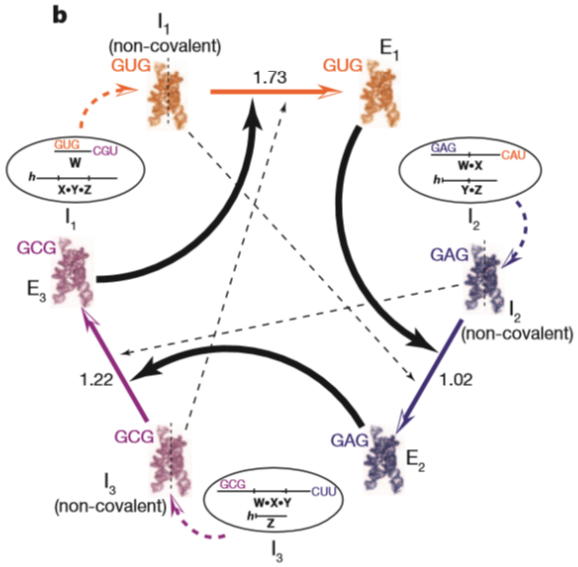

Fig. 1 The cooperative network is shown (selfish genotypes are not included). E1, E2 and E3 correspond to the 'UG', 'AA', and 'CU' genotypes respectively.

2 Model and Methods

The *Azoarcus* ribozyme is $\sim$ 200 nucleotides (nt) long, and can be broken into four fragments which covalently assemble by catalyzing recombination reactions[1]. The recombination reactions can occur with different efficiency determined by four of the 200 nucleotides in the sequence. One of these nucleotides defines the internal guiding system or **IGS**, while the other three serve as targets, or **tags**. Each tag is associated with a particu-

[a] *Address, Address, Town, Country. Fax: XX XXXX XXXX; Tel: XX XXXX XXXX; E-mail: xxxx@aaa.bbb.ccc*
[b] *Address, Address, Town, Country.*

IGS / TAG	U	C	G	A
U	0.0022	0.0038	0.0049	0.0197
C	0.0004	0.0016	0.0415	0.0020
G	0.0091	0.0125	0.0006	0.0005
A	0.0319	0.0069	0.0001	0.0004

Table 1 The catalytic rate constants for different IGS/TAG pairs are shown. The IGS are indexed on the vertical, while the tags are indexed horizontally.

lar 'tail' which identifies a break-point where the sequence was fragmented. By pairing a tag with non-complementary IGS a network of ribozymes can be created where any individual ribozyme cannot (effectively) catalyze it's own production but rather each can catalyze another recombination reaction. In this way, one can construct a cooperative cycle, such as the one shown in figure 1.

Following the empirical studies, we use 'W', 'X', 'Y' and 'Z', to label the ribozyme segments. In complete combinations of these fragments will be called fragments, while the complete sequence 'WXYZ' will be refereed to as an assembly or molecule [1]. The IGS is always associated with the 'W' segments, while 'W', 'X', and 'Y' may contain tags. 'Z' is the only fragment which never has an IGS nor a tag. A genotype can be understood in this context, as a combination as a specific combination of IGS and tag. In this numerical model we consider the following fragmented pairs, 1) 'WXY' and 'Z', 2) 'WX' and 'YZ' as well as 3) 'W' and 'XYZ.' The 'Y', 'X', and 'W' segments contain the tags, respectively. The different genotypes can be labeled as 'UG,' 'AA,' 'CU,' 'UA, 'AU,' and 'CG.' in accordance with figure (1). The cooperative genotypes are 'UG,' 'AA,' and 'CU,' while the selfish ones are 'UA, 'AU,' and 'CG.'

A kinetic Monte-Carlo algorithm was used to simulate the chemical kinetics [6]. For each experiment, all assemblies were initialized as fragments, with each genotype represented equally such that all fragments have an equal initial abundance, which was set to 500. This system is closed, therefore, for all time the mass of this system is fixed. A spontaneous formation rate constant was set to $k_s = 10^{-\eta}$. In this study, η was constant for any individual experiment, and it is assumed $\eta = 5$ unless otherwise stated. Empirical studies have estimated the catalytic efficiency of all IGS, tag pairings [2], those estimates can be found in Table 1. This numerical model is completely specified by the catalytic rate constants (table 1), the spontaneous formation rate constant (η), and the initial abundance of fragments.

The information measures used in this study are *Active Information* and *Transfer Entropy*. In the context of Shannon's Information Theory, Active information can be understood as reduction in uncertainty in a variable's future due to that variable's past k states. Active information is calculated using equation 1. Transfer Entropy can be understood as the reduction in uncertainty in the future state of one variable, due to the current state of another variable, conditioned on the previous k states of the first variable. Transfer Entropy is calculated using equation 2. These information measures were chosen specifically to compare the of information processing in this chemical system to the information processing in some biological systems [3].

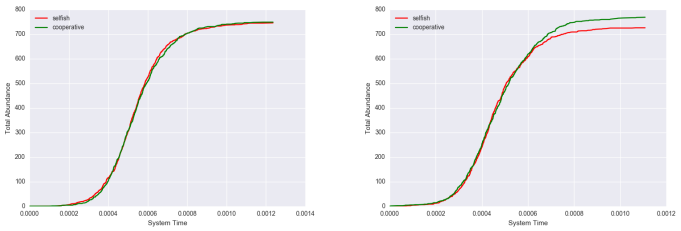

Fig. 2 The dynamics of the system are shown here. In green the total number of covalently assembled cooperative molecules are shown while the total number of the covalently assembled selfish molecules are shown in red.

$$AI_X^k = \sum_{(x_n^k, x_{n+1})} p(x_n^k, x_{n+1}) \log_2 \left(\frac{p(x_{n+1}, x_n^k)}{p(x_{n+1}) p(x_n^k)} \right) \quad (1)$$

$$T_{Y \to X}^k = \sum_{(x_n^k, x_{n+1}, y_n)} p(x_n^k, x_{n+1}, y_n) \log_2 \left(\frac{p(x_{n+1} | x_n^k, y_n)}{p(x_{n+1} | x_n^k)} \right) \quad (2)$$

The quantities of interest in this system are the abundances of each of the assembled genotypes. These quantities allow us to determine whether or not the cooperative cycle or the selfish individuals are selected. There are six distinct genotypes, each of which can take on real positive integer values, on the interval $[0, 500]$ (because the initial number of fragments is 500). In general, it will not be tractable to calculate 500^6 different probabilities, to determine the probability distributions used in equations (1) and (2). Therefore it is necessary to *coarse grain* the system dynamics. Here we coarse grain each iteration of the algorithm by ranking each genotype according to its abundance at that time. This effectively reduces the size of the state space to 6^6, rather than 500^6, reducing the number of probabilities by a factor of $\sim 10^{11}$. For the results shown here, the probability distributions were calculated using 10 different realizations of the system. For details on calculating probability distributions, please see [3].

3 Results

The kinetic model simulated here recreates the functional form of the growth laws observed in the empirical experiments [1]. Rather than simulating only the exponential growth phase, as was done in previous studies [1], this system models the long term growth of the molecules, as seen in figure 2. However, in contrast to the results presented in [1], it should be noted that the cooperative molecules do not always out compete the selfish ones.

The information measures tested were unable to distinguish the cooperative molecules from the selfish ones. Intuition would suggest that cooperative assemblies would be less determined by their own state, as they depend on the concentrations of other assemblies for their production. Naively this should correspond to lower values of Active Information for cooperative molecules relative to selfish molecules. The rank ordered plot 3 shows the active information for each individual genotype. The cooperative

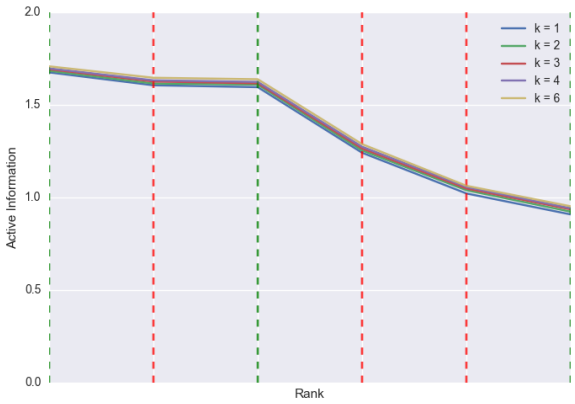

Fig. 3 Here the rank ordered Active Information is plotted for each genotype. The genotypes in the cooperative network are indicated by a dashed green line, while the selfish genotypes are shown with a dashed red line. History lengths (k) of $1,2,3,4,5,$and 6 are shown for $\eta = 3$.

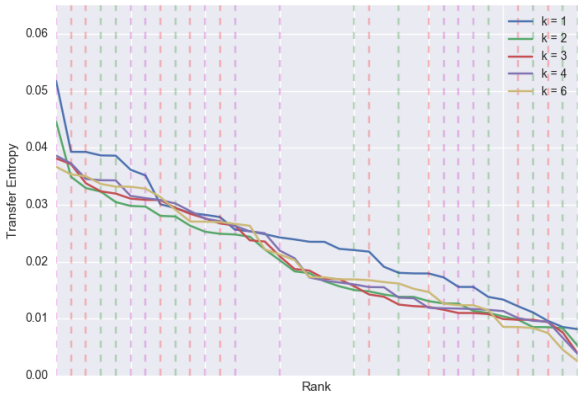

Fig. 4 The rank ordered Transfer Entropy between all pairs of genotypes are shown. History lengths (k) of $1,2,3,4,5,$and 6 are shown for $\eta = 3$. A history length of $k = 1$ is the most predictive. The vertical lines identify the relationships between nodes. The green lines correspond to interactions between cooperative molecules, the red lines correspond to interactions from a cooperative molecules to a selfish molecule, while purple lines correspond to interactions from selfish to cooperative molecules. The nearly linear scaling relation should be contrasted with the biological scaling relation found in [3]

genotypes are highlighted with a vertical green line while the selfish ones are highlighted with a vertical red line. There is no clear correlation between the active information processes and the nature of the genotype.

The transfer entropy scaling relation for this chemical system (shown in figure 4) is clearly different in nature from the one discovered in [3]. The biological scaling relation in [3] has multiple distinct features, while the chemical one in this study is nearly linear. In contrast to the biological network, there are no node pairs which have zero values for transfer entropy, and there are not plateus of transfer entropy, which can be seen in the biological scaling relation. The history length of $k = 1$ was the most predictive in this system. There is no clear relationship between a node pairs presence in the cooperative network and the transfer entropy between them. The transfer entropy calculated between pairs of genotypes does not appear to be correlated with the kinetic rate constants estimated by [2]. In figure 5 the transfer entropy between sources (vertical axis) and targets (horizontal axis) is shown. The most correlated genotypes are the 'UG' and 'CG' genotypes.

4 Discussion

The analyses here failed to find significant correlations between the system dynamics and the information measures. This is evidenced by the fact that the active information measure was unable to identify the cooperative molecules relative to the selfish ones (figure 3). The source of this failure could be due to one of two distinct factors. First, the coarse-grainning chosen here may not capture the relevant system dynamics. If this is the case, a new coarse grainning of the system dynamics might accurately identify the relevant features. Alternatively, it is possible that the information measures chosen are ill-suited to measure the types of correlations relevant to the system dynamics.

In order to test against the first case, alternative coarse grainning should be chosen and tested. An exhaustive search over all possible coarse grainning is not tractable. However more input

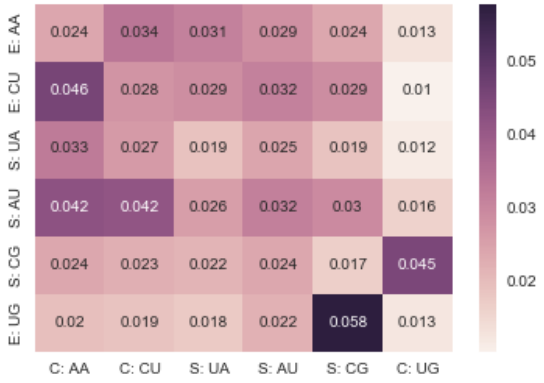

Fig. 5 Here the transfer entropy between every pair of genotype is shown. The 'source' (Y) is index on the vertical axis while the the target (X), is indexed on the horizontal axis. Darker colors correspond to more transfer entropy between pairs of genotypes. Selfish genotypes are identified with an S prefix while cooperative genotypes are identified with an E prefix

from empirical data may help identify viable representations. On the other hand, it may be necessary to compare different information measures in order to elucidate the system dynamics. In particular it may be important to consider global scale variables. Here only local variables were compared with other local variables. The dynamics here depend on a shared global resource (fragments with no tags) and therefore a comparison between global states and local states might provide more insight. Regardless of the interpretation, information theory provides a coherent way of comparing biological and chemical evolution. In this light, the use the study of information dynamics should be extended in the study of both biological and chemical evolution. This would allow more direct comparisons between theory of experiments and may provide new insights into the origin of life on earth.

References

1 N. Vaidya, M. L. Manapat, I. A. Chen, R. Xulvi-Brunet, E. J. Hayden and N. Lehman, *Nature*, 2012, **491**, 72–77.
2 J. A. Yeates, C. Hilbe, M. Zwick, M. A. Nowak and N. Lehman, *Proceedings of the National Academy of Sciences*, 2016, 201525273.
3 H. Kim, P. Davies and S. I. Walker, *Journal of The Royal Society Interface*, 2015, **12**, 20150944.
4 A. Wochner, J. Attwater, A. Coulson and P. Holliger, *Science*, 2011, **332**, 209–212.
5 H. S. Zaher and P. J. Unrau, *Rna*, 2007, **13**, 1017–1026.
6 D. T. Gillespie, *Journal of computational physics*, 1976, **22**, 403–434.